Lecture 14: DRAM, PCM

- Today: DRAM scheduling, reliability, PCM
- Class projects

TCM

Kim et al., MICRO 2010

- Organize threads into latency-sensitive ad bw-sensitive clusters based on memory intensity; former gets higher priority

- Within bw-sensitive cluster, priority is based on rank

- Rank is determined based on “niceness” of a thread and the rank is periodically shuffled with insertion shuffling or random shuffling (the former is used if there is a big gap in niceness)

- Threads with low row buffer hit rates and high bank level parallelism are considered “nice” to others

ECC

- For a BCH code, to correct t errors in k-bit data, need an r-bit code, $r = t * \text{ceil}(\log_2 k) + 1$

- For DRAM, typically, an 8-bit ECC (Hamming) code is attached to every 64-bit word; can recover from a single bit corruption

- Chipkill correct systems can withstand failure of an entire DRAM chip

- For chipkill correctness
  - the 72-bit word must be spread across 72 DRAM chips
  - or, a 13-bit word (8-bit data and 5-bit ECC) must be spread across 13 DRAM chips